DOCUMENT RESUME

ED 283 826 TM 870 277

AUTHOR Weiss, John
TITLE Truth-in-Testing & the Golden Rule Principle: Two Practical Reforms.
INSTITUTION National Center for Fair and Open Testing (FairTest), Cambridge, MA.
PUB DATE Apr 87
NOTE 17p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Washington, DC, April 21-23, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Access to Information; College Entrance Examinations; Confidentiality; Culture Fair Tests; Educational Testing; Higher Education; Information Needs; Licensing Examinations (Professions); Standardized Tests; *State Legislation; *Test Bias; *Test Construction; *Testing Problems; Test Validity
IDENTIFIERS *Golden Rule (Testing); National Teacher Examinations; Scholastic Aptitude Test; Test Publishers; *Truth in Testing

ABSTRACT

Two reforms are suggested as ways to improve standardized testing. They are Truth-in-Testing legislation and the Golden Rule bias reduction principle. Test publishers, state departments of education, and colleges who administer tests are often reluctant to provide statistical data on test results. It is also difficult for researchers to see test items. Truth-in-Testing legislation was enacted in California and New York to provide more information about college entrance examinations such as the Scholastic Aptitude Test. As a result, test items may be studied and challenged regarding their accuracy, validity, and cultural bias. The Golden Rule procedure is based on an out-of-court agreement between the Educational Testing Service, the State of Illinois and the Golden Rule Insurance Company concerning an insurance licensing examination. The agreement settled a lawsuit charging that ETS' Illinois Insurance Agent Licensing Exam unfairly discriminated against blacks and was not job-related. The National Center for Fair and Open Testing (FairTest) has endorsed the Golden Rule principle: "among questions of equal difficulty and validity in each content area, questions which display the least differences in passing rates between majority and minority test-takers should be used first." Safeguards need to be established to ensure that standardized tests measure relevant knowledge differences between test-takers and not irrelevant, culturally specific factors. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED283826

FairTest

National Center for Fair & Open Testing

TRUTH-IN-TESTING & THE GOLDEN RULE PRINCIPLE: TWO PRACTICAL REFORMS

Presented at a National Council of Measurement in Education's Invited Symposium "FAIRNESS IN TESTING: HAS THE INDUSTRY DONE ENOUGH?", April 22, 1987 in Washington, D.C.

by John Weiss

John Weiss, Executive Director of the National Center For Fair & Open Testing (FairTest), helped Congressman Michael J. Harrington draft the first federal Truth-in-Testing legislation in 1977.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Weiss

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM 870 277

BEST COPY AVAILABLE

Public policy protects the food we eat, the products we buy, the safety standards at our workplaces. Yet, despite its overwhelming impact on both citizens and institutions, America's multi-billion dollar testing industry is virtually unregulated. Few consumer protection laws apply, and test publishers refuse to voluntarily provide elected government officials, independent researchers, or test-takers with information needed to verify that their exams are fair or valid.

Even federal government sources of information on tests are controlled by test publishers. For example, America's largest testing firm, Educational Testing Service, operates the federal government's Education Research Information Center on Tests, Measurement and Evaluation (ERIC/TME), the nation's "official" clearinghouse distributing information on standardized tests and other methods of evaluating students and workers. This clearinghouse blatantly promotes ETS' products and philosophy. (1)

Both individuals who believe they have been wronged by a standardized test and the growing evaluation reform movement in this country have been seriously hampered by the inaccessibility of information. As the late Dr. Oscar Buros, editor of the authoritative Mental Measurements Yearbook, lamented:

> "today it is practically impossible for a competent test technician or test consumer to make a thorough appraisal of the construction, validation, and use of standardized tests...because of the limited amount of trustworthy information supplied by test publishers." (2)

Without access to data, individuals concerned about the misuses of tests cannot investigate alleged abuses or propose remedial procedures. Today, even raw statistics about the pass rates of different population groups are often difficult to obtain. As University of Florida Professor Pritchy Smith explained:

> "A Catch-22 scenario is often presented to the researcher. Testing companies most often insist that state departments of education have the responsibility for reporting test results. State departments, in turn, insist that since the testing company administers the test, it should report the results. The ultimate machination occurs, however, when both testing companies and state departments declare that it is not they but the colleges and universities that have the responsibility to report test results." (3)

It is equally difficult for independent researchers to gain access to exam questions. George Madaus, past President of the National Council on Measurement in Education, reports that "the simple truth is that many tests prepared by vendors for state agencies are secret, and not readily available for peer evaluation." (4)

Reform #1: Truth-in-Testing

In the 1970's a coalition of students, parents, teachers, principals, civil rights activists, and frustrated researchers joined together to press for the release of basic data about standardized tests. Over the strenuous opposition of the testing industry, this coalition convinced legislators in

California and New York State to enact Truth-in-Testing laws for college and graduate school entrance exams. (5)

These Truth-in-Testing laws enable students taking university admissions tests to obtain copies of the questions used to determine their scores, their answer sheets and the "correct" answer key. Such procedures enable test-takers to learn from their mistakes, to provide future applicants with a better understanding of the areas covered in the tests and allow independent researchers to examine test questions for fairness and accuracy. (6) These laws also forced test publishers to make public studies concerning the accuracy, accuracy, validity and cultural biases of their tests. In addition, as a result of these initiatives, test publishers have been forced to publicly defend or re-score items which test-takers have challenged as inaccurate or biased. More than 500,000 students scores have been raised after students discovered errors on university admissions tests through Truth-in-Testing laws. (7)

More disturbing than the ambiguous or erroneous questions has been the discovery that hundreds of items on these academic admissions exams either require knowledge of the activities and vocabulary of upper-middle class Americans or have a distractor answer that often fools minority and/or female test-takers. (8) Recent SAT items disclosed due to Truth-in-Testing expected students to be familiar with polo, golfing, tennis, pirouettes, minuets, property taxes, violins, melodeons, tympanists, and horsebackriding. (See Exhibit 1.) Students who lack this kind of culturally specific knowledge cannot obtain the high SAT

Exhibit 1

Scholastic Aptitude Test Questions

The following SAT items represent a small portion of the university admission test questions that FairTest claims require knowledge of the activities or vocabulary of upper-middle class America.

1) **RACKET:TENNIS**

A) springboard:diver
B) horse:polo
C) glove:boxing
D) club:golf
E) gun:hunting

2) **HEIRLOOM:INHERITANCE**

A) payment:currency
B) belongings:receipt
C) land:construction
D) legacy:bill
E) booty:plunder

3) **MINUET:DANCE**

A) beret:bowler
B) clarinet:symphony
C) chariot:wheel
D) sonnet:poem
E) gown:petticoat

4) **BRIDLE:HORSE**

A) bone:dog
B) olive branch:dove
C) valor:soldier
D) precept:conduct
E) devotion:duty

5) **MELODEON:ORGANIST**

A) reveille:bugler
B) solo:accompanist
C) crescendo:pianist
D) anthem:choirmaster
E) kettledrum:tympanist

wanted answers: 1) D 2) E 3) D 4) D 5) E

needed to enter most of America's selective colleges and receive financial aid awards from many private foundations and government agencies. (9) Examine the following SAT item:

RUNNER:MARATHON

(A) envoy:embassy

(B) martyr:massacre

(C) oarsman:regatta

(D) referee:tournament

(E) horse:stable

Fifty-three percent of the whites, but just 22% of the blacks gave the wanted answer (C). (10) Clearly this item does not measure students' "aptitude" or logical reasoning ability, but instead knowledge of an upper-middle class recreational activity.

ETS has acknowledged that information made public through Truth-in-Testing laws has caused them to institute new procedures designed to detect and remove unfair items from their tests. (11) Unfortunately, current Truth-in-Testing laws only cover university admissions tests, which constitute fewer than 20% of the standardized exams administered each year. (12)

If Truth-in-Testing policies are to be extended to additional tests, special provisions need to be made for low volume exams. When test questions are disclosed they can not be reused. For high volume tests, such as the SAT, the cost of developing new test forms is minimal. (13) For low-volume tests, the cost on a

per candidate basis would be very large. Truth-in-Testing legislation for professional licensing exams currently pending in New York State uses a sliding scale for disclosure. High volume tests must be disclosed after every administration. For low volume tests (annually administered to less than 5,000 individuals), just one form must be disclosed every five years. (14)

The New York legislature is also debating a bill which would create a special commission to examine New York State's use of university admissions tests. In addition, the legislation requires test companies to produce a gender and racial "item analysis" for every scored item on their exams. With such data, independent researchers will be able to determine which questions have unusual scoring patterns. Then these questions can be investigated to determine whether the question was appropriate, or if it was actually measuring irrelevant or culturally-specific knowledge. (15)

Reform #2: The Golden Rule Bias Reduction Principle

A decade ago, Educational Testing Service researchers Donald Medley and Thomas Quirk examined ETS' National Teachers Exam (NTE). Their goal was to determine whether the NTE unfairly discriminated against black test-takers. They undertook this research after noticing that:

> "a reading of the items on a typical form of the [NTE] would have revealed little awareness on the part of the test's constructors of the fact that this country contains a large minority of black citizens represented by many fine writers, artists and musicians..." (16)

Professors Medley and Quirk had ETS' test developers prepare an experimental section of the NTE which was administered to NTE test-takers as a non-scored section. Test-takers did not know that this section would not be scored. One-third of the items on this experimental section:

> "met the usual specifications as to content; one-third met the same specifications but reflected black culture rather than traditional culture represented by the first group; and one-third of reflected modern (non-black) culture rather than traditional." (17)
>
> "The "black' and "modern' items were not esoteric in the sense that they called for knowledge peculiar to one group; rather, the Black and modern items dealt with material any reasonably well-informed person might be expected to know, but which happened to reflect the experiences, accomplishments, and concerns of black citizens (in the one case) and of citizens who lived in the last few decades (in the other)...Each experimental item underwent the normal procedures of construction and review maintained for NTE items." (18)

Medley and Quirk then examined the results of Black and white NTE test-takers who took their experimental test, controlling for sex and urban/rural differences between test-takers. They found that the average black test-taker scored 4.4 points <u>higher</u> than their white counterpart on the "black" items, and 13.4 points <u>lower</u> on the "traditional" items.

The two ETS researchers then concluded:

> "Clearly, whatever else this test may measure, it has the potential of measuring a candidate's racial background with considerable accuracy...The discovery that an unsuspected factor, such as whether a personality referred to in a test is black or white, may have more to do with candidates' scores than manifest content...is disturbing." (19)

Evidence recently made public through court cases in Alabama (20), California (21), Illinois (22), New York (23), and Texas (24) demonstrates that some testing companies--and the state agencies which authorize their activities--have not taken sufficient care to eliminate biased items from their exams.

Safeguards need to be established to ensure that standardized tests measure relevant knowledge differences between test-takers, and not irrelevant, culturally-specific factors. One practical procedure currently available to make tests as fair as possible is the Golden Rule procedure. This is an objective technique, based on a November, 1984 out-of-court agreement between the Educational Testing Service, the State of Illinois, and the Golden Rule Insurance Company. (25) The agreement settled a lawsuit charging that ETS' Illinois Insurance Agent Licensing Exam unfairly discriminated against blacks and was not job-related.

Under Golden Rule, the same content areas will be covered as on previous tests and the exam will be of the same overall level of difficulty. The only difference is that within groups of equally valid items in the same content areas, test

publishers must select those items which display the smallest differences between the correct answer rates of minority and majority test-takers. As Emory University Professor Martin Shapiro told the New York Times, "Once you have this method, not to use it is to knowingly use a more discriminatory procedure." (26)

To understand why the Golden Rule reform is needed, it is necessary to examine how test-makers currently construct exams, including teacher licensing tests. For each content area, test publishers develop a pool of potential questions. They then pre-test these questions on a sample group of test-takers. Next, test publishers discard those items which they believe are ambiguous, biased, or otherwise flawed. From the remaining pool of items they employ a statistical technique--based on point-biserial correlations--to select for the final test those pre-tested questions which maximize the differences between test-takers. Questions which maximize differences between high and low scoring students may really be measuring test-takers' knowledge of irrelevant, culturally-specific information. By using such items, tests may discriminate against otherwise qualified individuals.

The Golden Rule reform makes exams fairer, not easier. Nothing in Golden Rule changes the content specification of future exams. Nothing alters the ability of test administrators to set cut scores or passing rates. If group performances differ widely on all the items in a given content area, these questions will still be used on exams assembled under Golden Rule.

Setting the Record Straight

In recent months, several individuals have misinterpreted FairTest's position on Golden Rule. To set the record straight, FairTest does not believe, as ETS President Greg Anrig asserts, "that group differences on test questions primarily are caused by 'bias'." (27) Rather, we recognize that group score differences reflect a host of causes, including genuine knowledge differences, test-taking abilities, as well as the inclusion on tests of irrelevant and biased questions. The purpose of the Golden Rule reform is to help ensure that biased test questions are removed from exams.

Application of the Golden Rule settlement to ETS' Illinois Insurance Licensing Test has had just this impact. According to data revealed to the Advisory Committee established to monitor the agreement's implementation, the gap between the average scores of black and white test-takers closed by twenty-five percent under the first Golden Rule assembled forms. (28) If many minorities score just below the passing cut-off, a 25% reduction in the disparate impact of the exam will lead to a substantial increase in the number of minority candidates who pass--without changing the exam's validity or making it easier.

Another misinterpretation of FairTest's position on Golden Rule, made recently by University of Illinois Professor Robert Linn and National Evaluation System's Attorney Michael Rubell, is that we want to simply extend the out-of-court settlement reached in Illinois. (29) FairTest has always recognized that the Illinois settlement has many special provisions that were

either incorporated into the settlement to appease one of the parties in the case or are only appropriate for an insurance licensing exam. FairTest has only endorsed the Golden Rule principle, which is that:

> among questions of equal difficulty and validity in each content area, questions which display the least differences in passing rates between majority and minority test-takers should be used first. (30)

* * * * * *

Standardized multiple-choice exams have become our nation's cradle-to-grave arbiter of social mobility. These exams are far too important to be left to the sole control of those who profit from their sales. The modest, practical proposals outlined above are needed first steps in what is sure to be a very long campaign to ensure that the 40 million standardized multiple-choice tests annually administered to America's students and job applicants are fair, open and pedagogically sound.

FOOTNOTES

1) See Chronicle of Higher Education, September 10, 1986, and Education Week, November 19, 1986. Readers interested in obtaining documentation of FairTest's charges that ETS "blatantly" uses the the ERIC/TME clearinghouse to promote its products and philosophy" should send a self-addressed stamped envelope (39 cents postage) to FairTest, Box 1272, Harvard Square Station, Cambridge MA 02238.

2. Quoted in Andrew J. Strenio, Jr., The Testing Trap, , Rawson, Wade Publisher, Inc. 1981, p. 274.

3. G. Pritchy Smith, "Unresolved Issues and New Developments in Teacher Competency Testing," Urban Educator, Fall 1986.

4. George Madaus, "Public policy and the testing profession--you've never had it so good?" National Council of Measurement in Education Presidential Address, April 2, 1985.

5. See Chapter 1 of None of the Above by David Owen Houghton-Mifflin, 1985 (Boston) for a description of the testing industry's initial opposition to Truth-in-Testing legislation.

6) See "Teaching Knowledge: How Do We Test It" by Linda Darling-Hammond, American Educator Fall 1986 for examples of how tests can be ambiguous. Hammond concludes: "About 40 percent of the questions [on ETS' National Teachers Exam sample test], by my reckoning, did not have a "right" answer, either because the question contained insufficient information to allow a complete evaluation of the situation or because alternative answers would be correct depending upon what research one relied on."

7) From a National Center for Fair & Open Testing research study in progress.

8) See The GMAT Exposed & The LSAT Exposed by David M. White Testing For The Public, 1985 for examples of questions that contain distractor answers that often fool minority candidates. For examples of questions that contain distractor items that often fool females, contact National Center for Fair & Open Testing.

9) See "Sex Bias in College Admissions Tests: Why Women Lose Out", (National Center Foir Fair & Open Testing, April 1986) for a description of how low test scores reduce students abilities to obtain scholarships and gain admission to selective institutions.

10) Thomas F. Donlon, "The SAT in a Diverse Society: Fairness and Sensitivity," College Board News #122, Winter 1981-82.

11) "Nation's Test-Giver Offers Some Answers", March 23, 1986, The New York Times page E 24.

12) Statistics from a National Center for Fair & Open Testing research study in progress.

13) An internal ETS "Activity Analysis" made public by Allan Nairn & Ralph Nader documents that in 1972 it cost ETS just 33 cents per candidate to develop all the questions used on an SAT form. See The Reign of ETS by Allan Nairn & Ralph Nader, Center for Responsive Law, 1980. page 148-150.

14) This legislation is being co-sponsored by State Senator Kenneth LaValle and Deputy Speaker of the House Arthur Eve.

15) This legislation is being co-sponsored by State Senator Kenneth LaValle and Deputy Speaker of the House Arthur Eve.

16) Donald M. Medley and Thomas J. Quirk, "The Application of a factorial design to the study of cultural-bias in general culture items on the National Teacher Examination", Journal of Educational Measurement, Vol. II, No. 4, Winter 1974.

17) Ibid.

18) Ibid.

19) Ibid.

20) Margaret T. Allen v. Alabama Board of Education, 612 F. Supp. 1046 CMD Ala. 1985. See FairTest Examiner, Spring 1987.

21) Larry P. v. Riles. See FairTest Examiner, Spring 1987, p. 10.

22) Golden Rule Insurance Company v. Washburn No. 419-76 (Illinois Circuit Court, 7th Ind. Cir. Ct., November 20, 1984 Consent Decree.

23) Eastern District of New York, Civil Action #77, Civ. 1881 (FXA). Also see "Justice Dept. Criticizes Company in Preparing Exam for Long Island Police" New York Times, May 26, 1985 page 33.

24) UCAC v. State of Texas, 5th Circuit Court, Texas, Case No. 85-2579.

25) Golden Rule Insurance Company v. Washburn No. 419-76 (Illinois Circuit Court, 7th Ind. Cir. Ct., November 20, 1984 Consent Decree.

26) New York Times, November 29, 1984.

27) APA Monitor, January 1987. Also see FairTest's "Response to Anrig," APA Monitor, April 1987.

28) From Emory University Professor Martin Shapiro's presentation at FairTest's 1987 Convention, "Testing Reform Initiatives: 1987 & Beyond," December 11, 1986, Capitol Hill, Washington, D.C.

29) See Michael A. Rebell, "Disparate Impact of Teacher Competency Testing on Minorities," Yale Law & Policy Review, Spring/Summer 1986, and Greg Anrig, APA Monitor, January 1987. and Robert L. Linn and Fritz Drasgow, "Implications of the Golden Rule Agreement, The Score, APA Newsletter, January 1987. p. 4.

30. A Golden Rule Bias Reduction Procedure Sourcebook is available from FairTest for $20.00. P.O. Box 1272, Cambridge MA 02238.